

Data is the new Oli?

Development of a Metadata-Crawler for HPC Simulations

Part of good scientific practice is making ones research data FAIR, i.e. findable, accessible, interoperable and reusable. This is achieved by attaching rich metadata to the research data. The relevant metadata fields are described in ontologies. Filling these ontologies by hand, however, is infeasible. Therefore, in this work, the metadata extraction is to be automated using python. The tool should first extract metadata from runs of our in-house computational fluid dynamics code and then successively extend to data from the the HPC system used and post-processing algorithms. Therein, flexibility to work on inputs from other sources, e.g. other codes or experimental data is to be ensured.

Tasks:

- Familiarization with ALPACA¹
- Familiarization with LRZ Linux-Cluster² and Slurm job scheduling³
- Familiarization with ontologies⁴
- Development of a python toolkit to extract metadata and fill respective ontology

Requirements:

- Knowledge of python
- Ability to work independently
- Knowledge of Linux-CL
- Experience with HPC-clusters, beneficial

Take-away:

- Experience with state-of-the art metadata concepts
- Experience with HPC-clusters
- Project management skills
- Improved Linux and python skills

1 <https://gitlab.lrz.de/nanoshock/ALPACA>
2 <https://doku.lrz.de/display/PUBLIC/Linux+Cluster>
3 <https://slurm.schedmd.com/>
4 [https://en.wikipedia.org/wiki/Ontology_\(information_science\)](https://en.wikipedia.org/wiki/Ontology_(information_science))

Contact:

Nils Hoppe
Room MW1617
E-Mail nils.hoppe@tum.de
Phone 089/289-16336

